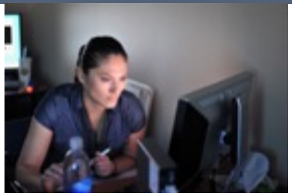
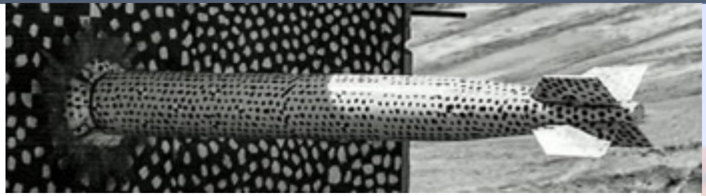
Sandia National Laboratories

Coupling of Novel Probabilistic Transfer Learning Strategies and Autoencoders to Expedite Turbulent Combustion Modeling


MMLDT-CSET 2021

9/28/2021

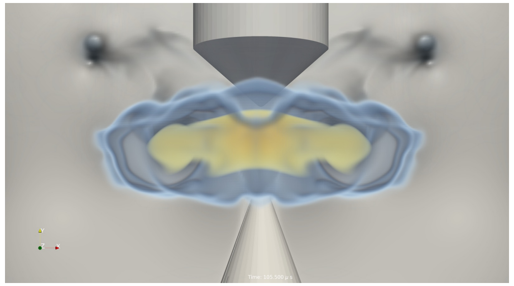
B. S. Soriano[1], Y. Shimizu[1], M. Rieth[1], M. Khalil[1], J. Chen[1], T. Echekki[2]

[1] Sandia National Laboratories, Livermore, CA

[2] North Carolina State University, Raleigh, United States

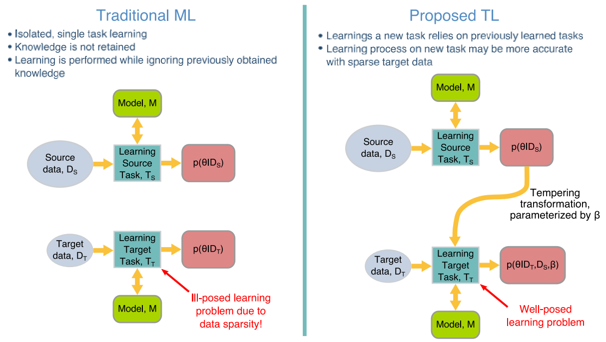
Summary

- Combustion science relies on Direct Numerical Simulation (DNS)
- DNS of reactive flows is expensive : ~10 billion DOF and ~10 million CPU-hours for relevant (3D) problem
- The key phenomena are often occurring in a lower dimensional manifold



Principal Component Analysis (PCA) is an alternative to expedite DNS calculations

- Significant dimensionality reduction in chemical species space
- Requires knowledge of the composition space accessed by (expensive) simulations *a priori*

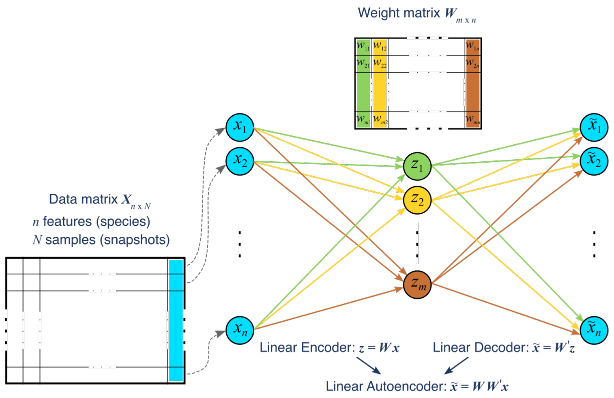


Transfer learning

- Leverage PCA models trained/calibrated for similar regimes of operation (source domains)
- Improve the training process on a target domain having limited data

Autoencoders for probabilistic characterization of PCA modes

- Utilize Autoencoders with linear activation functions to mimic dimensionality reduction achieved via PCA
- Autoencoders calibrated using **Bayesian methodologies** based on variational inference



Application of new probabilistic transfer learning framework

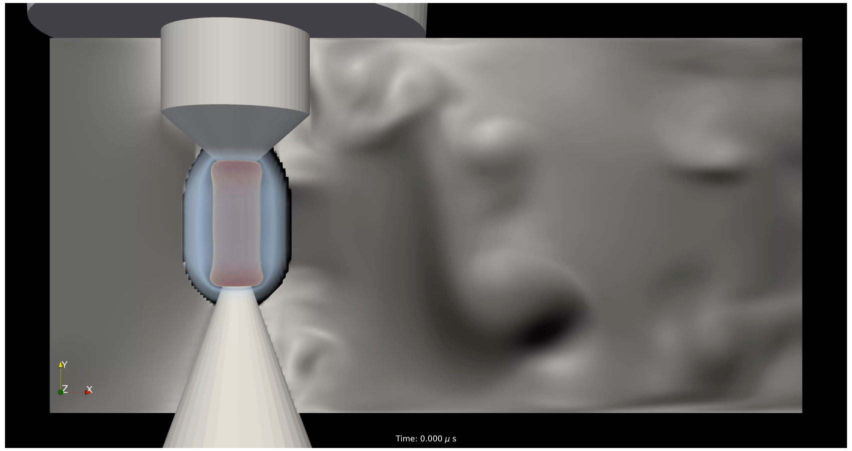
Introduce DNS for Turbulent Combustion

- DNS simulations often target conditions related to propulsion systems such as gas turbines and internal combustion engines
- Experimental measurements provided limited information about the highly non-linear problem
  - DNS is necessary to develop reduced-order models

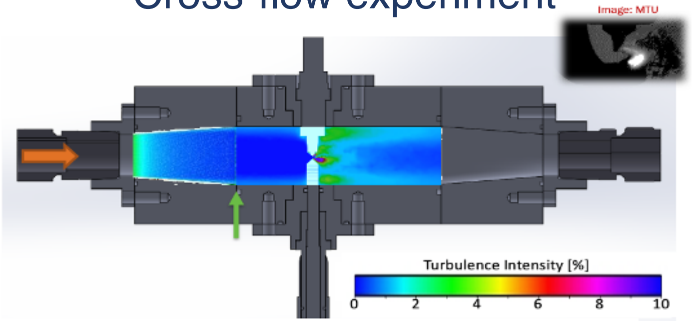
Exemplary DNS set up

- Pele code (part of the Exascale Computing Project - ECP): finite volumes formulation; Adaptive Mesh Refinement; Embedded Boundary treatment
- Approximately 1 billion DOF: with 98 chemical species
- Limitation: limited runs on parametric space

Spark-ignition DNS under cross-flow conditions



Cross-flow experiment



Courtesy: Isaac Ekoto (SNL)

Sandia National Laboratories

- PCs are optimum and generic parameters to describe the combustion state space manifold (in lieu of the traditional parameters, e.g. mixture fraction).
- PCA: a dimension reduction technique that converts a set of correlated variables (species and temperature) to weakly correlated ones:

Mathematically: $\boldsymbol{\phi} = \mathbf{Q}^T \boldsymbol{\theta}$

$\boldsymbol{\phi}$: PCs vector (size *N*), $\boldsymbol{\theta}$: representative species (size ***N***), $\boldsymbol{Q}$: matrix of eigenvectors of the covariance matrix of $\boldsymbol{\theta}$ (size $N \times N$)

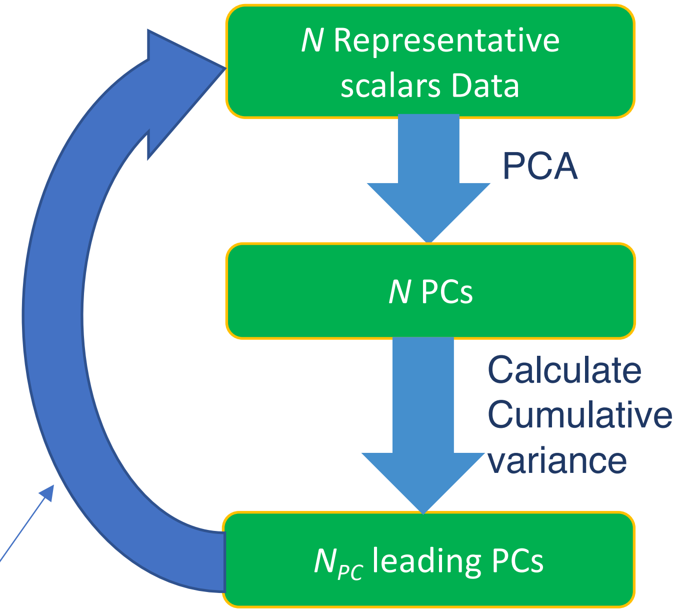
- Strategy: Retain a subset (N_{PC}<< *N*) of the PCs that represent the bulk of the data variance:

$$\boldsymbol{\phi}^{\mathbf{red}} = \mathbf{A}^T \boldsymbol{\theta}, \text{ with } \mathbf{A} \text{ the leading } N_{PC} \text{ vectors of } \mathbf{Q}$$

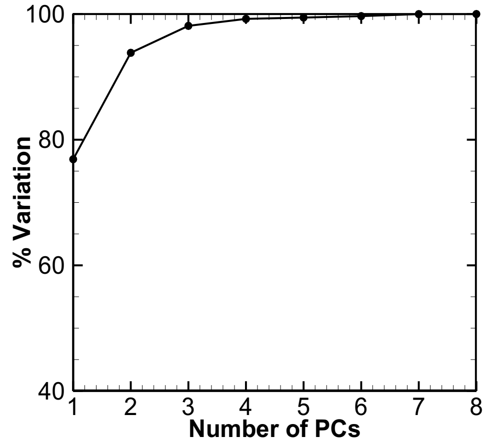
- Linear: thermo-chemical scalars to PCs; Recover non-linearly from PCs and thermo-chemical scalars (here using artificial neural networks, ANNs)
- Instantaneous transport equations for the PCs in DNS can be derived (Sutherland and Parente, 2009):

$$\frac{\partial \rho \phi_k}{\partial t} + \frac{\partial \rho u_j \phi_k}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\rho D_k \frac{\partial \phi_k}{\partial x_j}\right] + s_{\phi_k},\ k = 1, \cdots, N, \text{ where } \mathbf{s}_{\boldsymbol{\phi}} = \mathbf{A}^{\mathbf{T}} \mathbf{s}_{\boldsymbol{\theta}}$$

- These equations can be averaged (RANS) or filtered (LES)



Inversion from PCs to scalars using artificial neural networks

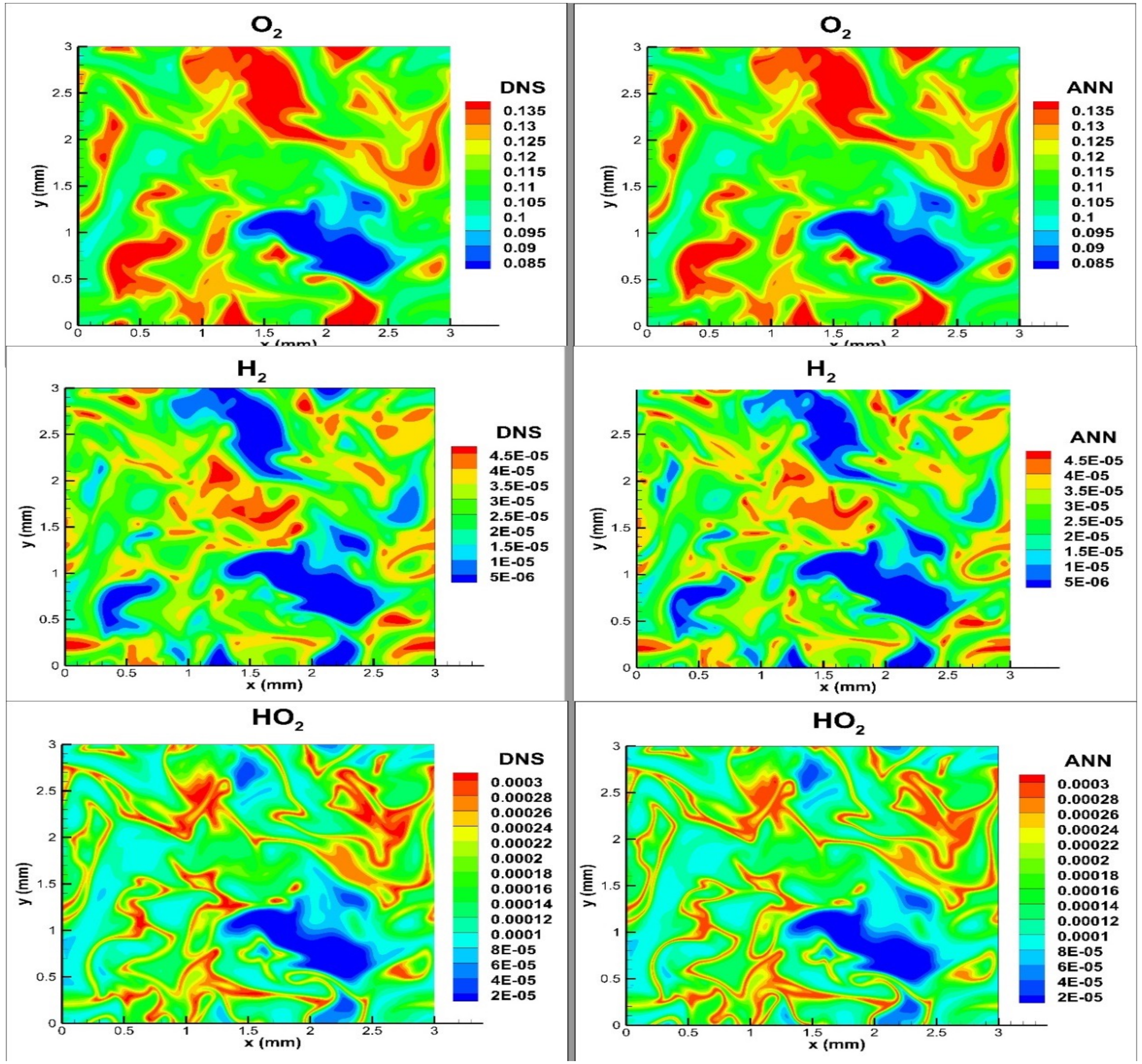


A Priori: 3D DNS of SACI with Ethanol (C2H5OH)

A Priori … (an important step in PC transport)

- Determine PCs from slides of solution
- Retain subset of the PCs
- Reconstruct solutions at different snapshots using the retained PCs.

- Original Manifold: 29 dimensional (28 species + temperature)
- Linear PCA performed in a subspace spanned of 6 representative scalars: T, O_2, H_2O, CO, CO_2 and C_2H_5OH .
- PCA-ANN tabulation of all 29 variables are satisfactory based on first 2 PCs.
- Potentially, an order of magnitude saving in computational time.
- Also potential for reducing stiffness if fast reactions are eliminated from the reduction process.

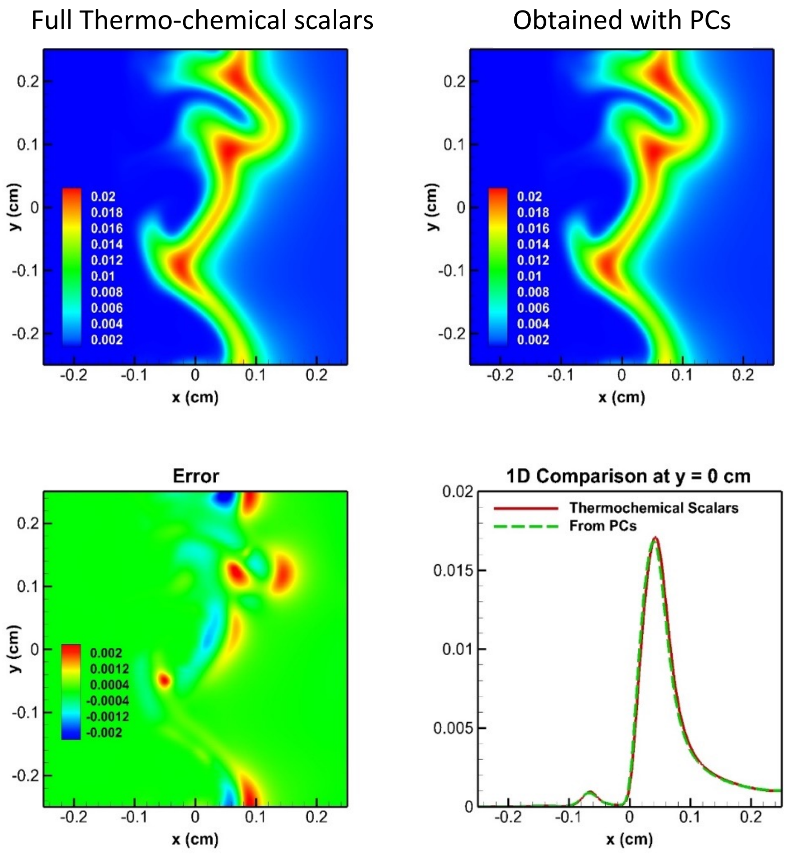


A Posteriori: 2D DNS of Methane-Air Premixed Flame wrinkling

***A Posteriori*: PC Transport in DNS (an *a priori* step is needed prior to *a posteriori*)**

- **(31) Thermo-chemical scalars**: (30 species + temperature) and 184 reactions
- **(8) Representative scalars**: T, CH_4, O_2, H_2O, CO_2 , CO, H_2 and O (O need to capture curvature/differential diffusion effects).
  - **(8) Potential PCs**
- **(4) Retained PCs**
  - The PCs capture the flame topology and are correlated with different key scalars.
- Saving in computational cost:
  - **4 vs. 31** scalars transported
  - A factor of **4** spatial resolution saving
  - A factor of **10** temporal resolution saving
- 2D DNS with species and energy has a similar computational cost to 3D DNS with PCs (huge saving)
- PC transport is not limited to a particular combustion mode.

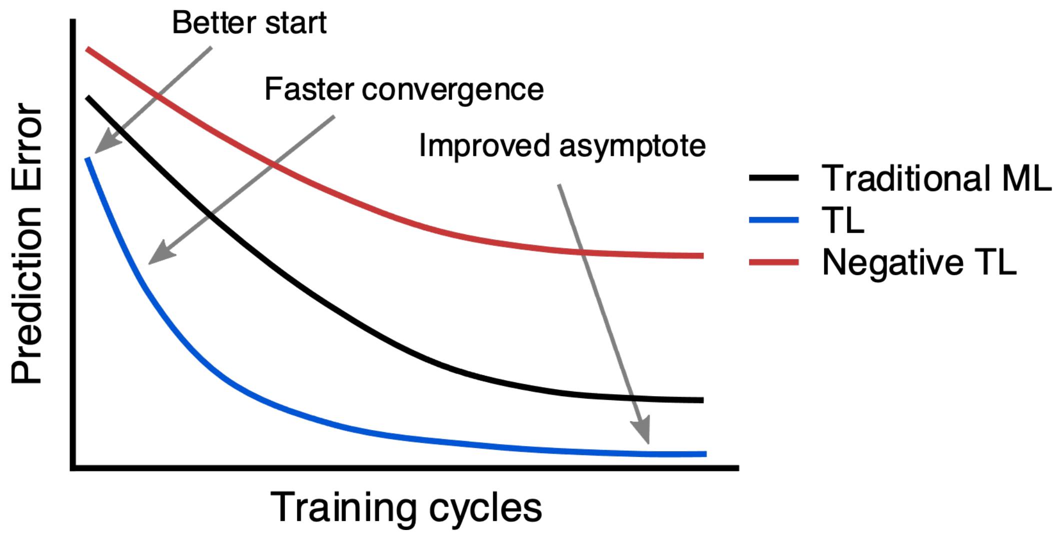
CO mass fraction



(Owoyele & Echekki, 2018)

Alleviating Ill-Posedness in Training of PCA Models in Sparse Data Settings

- **Challenge:** Training of PCA model requires knowledge of the composition space accessed by the simulations a priori
  - ✗ Excessive expense of DNS simulations limits the applicability of PCA due to lack of sufficient training data (snapshots)
- **Goal:** Aid the construction of PCA models predictions within sparse (and possibly noisy) data settings
- **Proposed Solution:** Application of novel probabilistic transfer learning framework
- Transfer learning (TL): knowledge gained through similar training tasks is used to possibly improve the training process on a target domain having limited/noisy data:
  - ✓ Improved initialization
  - ✓ Increased rate of convergence
  - ✓ Greater achievable performance
- Proposed framework will aim to alleviate potential negative transfer: TL resulting in decreased Performance

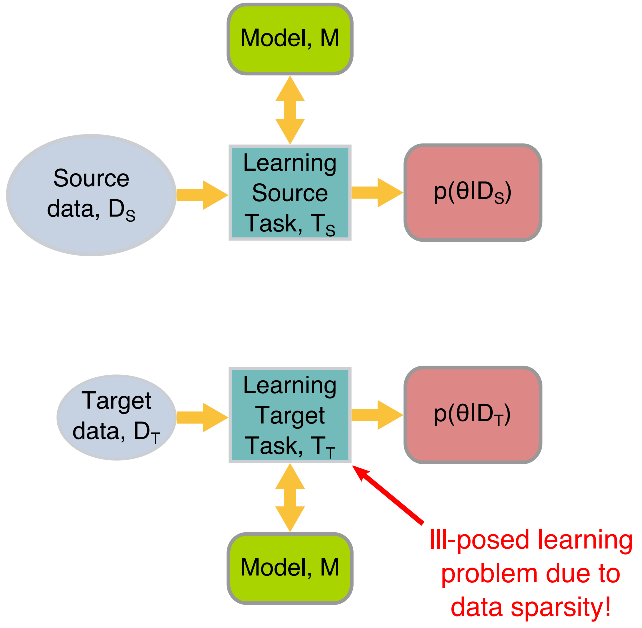


Novel Probabilistic Transfer Learning Framework

- Proposed TL framework aims to address the shortcomings in existing methodologies: It determines when to apply TL, which model to use, and how much knowledge to transfer.
- It relies on probability/measure theories to characterize and propagate uncertainties, thereby enhancing the trustworthiness of ML models in making predictions based on noisy and sparse training data.

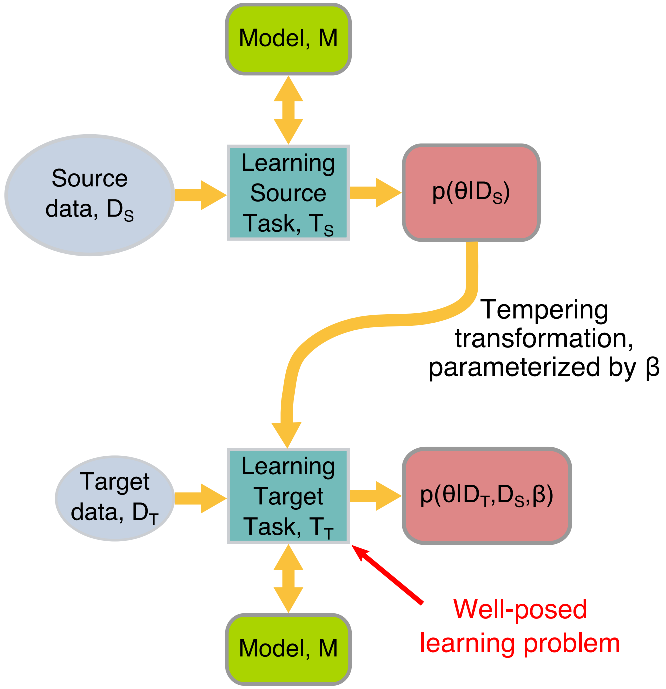
Traditional ML

- Isolated, single task learning
- Knowledge is not retained
- Learning is performed while ignoring previously obtained knowledge



Proposed TL

- Learnings a new task relies on previously learned tasks
- Learning process on new task may be more accurate with sparse target data



Sandia National Laboratories

$$\mathcal{M}(x,\theta) = y \approx d + \epsilon$$

ML model, features, parameters, target, observation, noise

- Forward Problem: Given ML model, $\mathcal{M}$, model parameters, θ, and feature vector, x, predict “clean” targets, y
- Inverse Problem: Given a set of “noisy” observations, $D = \{d_1, \dots, d_N\}$, and feature vectors, $X = \{x_1, \dots, x_N\}$, infer parameters
  - Observations are
    - inherently noisy with unknown (or weakly known) noise model
    - sparse in space and time (insufficient resolution)
  - Problem typically ill-posed, i.e. no guarantee of solution existence nor uniqueness
- Solution: Probability density function (PDF) over the parameter space obtained using Bayes’ rule:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

likelihood, prior, posterior, evidence

  - $p(\theta)$ is the prior PDF of θ: describes prior knowledge, inducing regularization
  - $p(d|\theta)$ is the likelihood PDF of θ: describes data fit
  - $p(\theta|d)$ is the posterior PDF of θ: full Bayesian solution
    - Not a single point estimate
    - Completely characterizes the uncertainty in θ
    - Subsequently used in making predictions under uncertainty

Transferal of "Learning" in a Probabilistic Setting

- In a transfer learning context, we have a target task of interest (regression/classification) with associated target data, D_T. We also have access to "supplementary" source data, D_S.
- **Our idea:** Extend mechanisms of propagating knowledge in sequential data assimilation (e.g. Kalman-based filters) by which captured knowledge from the source training task is used as prior knowledge in the target task:

$$\underbrace{p(\theta|D_T, D_S)}_{\text{posterior}} \propto \underbrace{p(D_T|\theta)}_{\text{likelihood of target data}} \underbrace{p_S(\theta)}_{\text{prior from source data}}$$

- **Challenge:** This approach does not provide flexibility in allowing the modeler to dictate how much knowledge, if any, is transferred from source to target tasks
- **Our solution:** Tempering-based methodologies
- The following is an example of the extension of power-based prior tempering transformation to "diffuse" knowledge in the prior PDF:

$$p(\theta|D,\beta) \propto p(D|\theta)p(\theta)^{\beta}$$

  - Full transfer: $\beta \rightarrow 1$ reverts back to the full likelihood from the source training task (i.e. traditional Bayes)
  - No transfer: $\beta \rightarrow 0$ results in a flat prior
  - Partial transfer: $0 < \beta < 1$

Degree of Knowledge/Learning to Transfer - Empirical Bayes

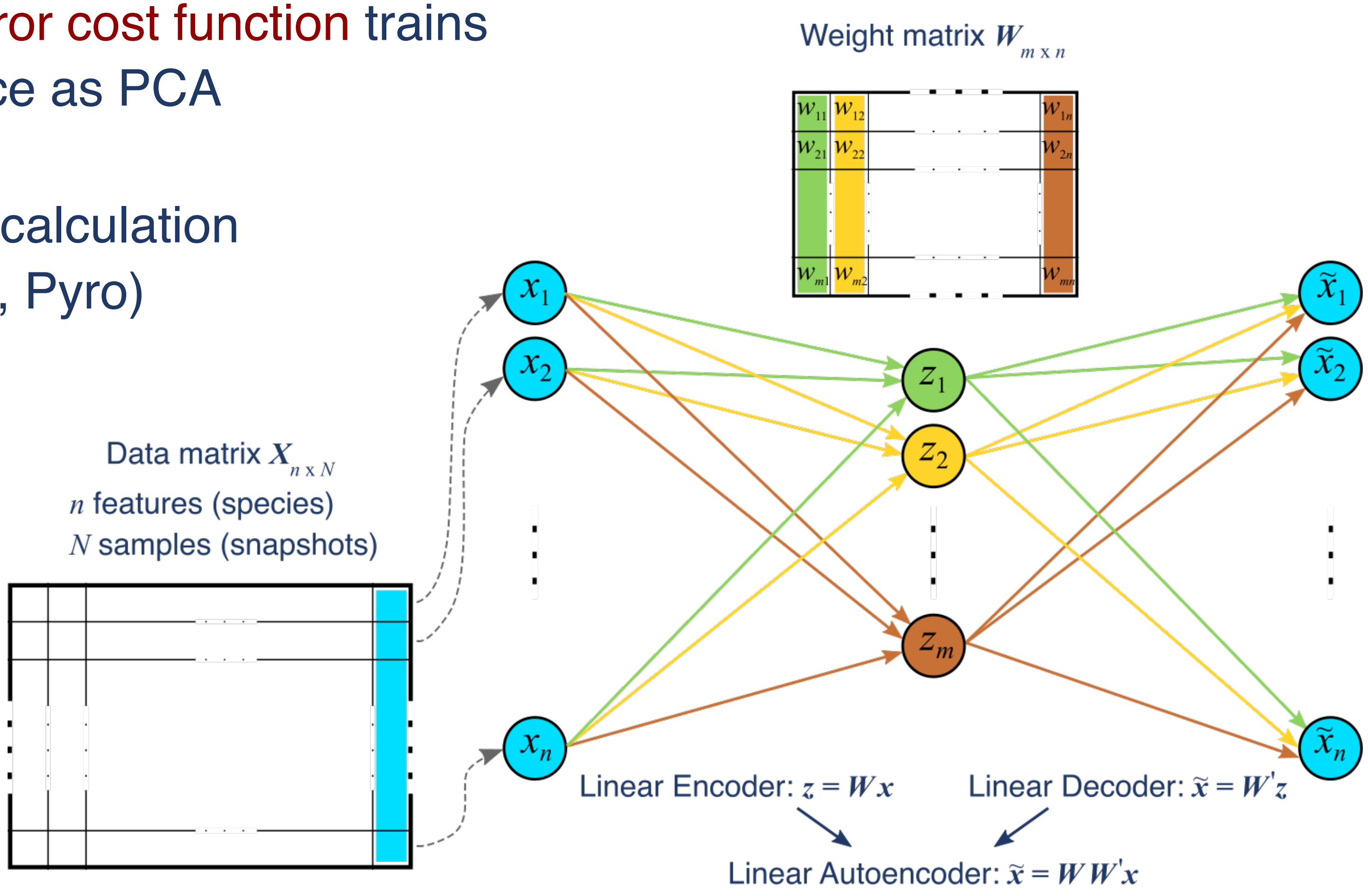
- The "tempering" hyper-parameter(s) β allow us to control the degree to which learning is transferred from the source task to the target task
- **Our idea:** Empirical Bayesian treatment to provide point estimates associated with ***some*** objective function
- **Challenge: What objective function?** Although empirical Bayes has been applied in numerous contexts for various purposes, there is not precedent for its use in transfer learning in determining such hyper-parameters
- **Our solution:** Follow an ***information-theoretic*** approach, focusing on the **relative entropy** that measures the information geometry in moving from the prior to posterior:

$$\underbrace{\mathrm{D}_{\mathrm{KL}}\left[\, p(\theta \mid D, \beta), p(\theta \mid \beta)\right]}_{\substack{\text{Relative entropy} \\ \text{(prior \& posterior)}}} = \underbrace{-\mathrm{H}\left[\, p(\theta \mid D, \beta), p(D \mid \theta)\right]}_{\substack{\text{Cross entropy} \\ \text{(likelihood \& posterior)}}} = \underbrace{\int \log p(D \mid \theta)\, p(\theta \mid D, \beta)}_{\substack{\text{Expected data-fit} \\ \text{(posterior-averaged log-likelihood)}}}$$

- This objective function (usually employed for Bayesian experimental design) has multiple interpretations, including an "average" log-likelihood one, resulting in a maximum "expected" likelihood estimate for hyper-parameter(s) β

Probabilistic PCA via Autoencoders and Variational Inference

- Challenge: The TL framework requires a probabilistic characterization of PCA modes
- Solution: Bayesian Autoencoders
  - Autoencoders with single fully-connected hidden layer, linear activation function and a squared error cost function trains weights that span the same subspace as PCA
- ✓ Trained with variational inference
- ✓ Leverage backpropagation for gradient calculation
- ✓ Leverage TPLs (TensorFlow Probability, Pyro)
- ✓ Easily switch to nonlinear activation functions (nonlinear-PCA)
- ✓ PCA modes can be recovered from autoencoder weights [Plaut, 2018]

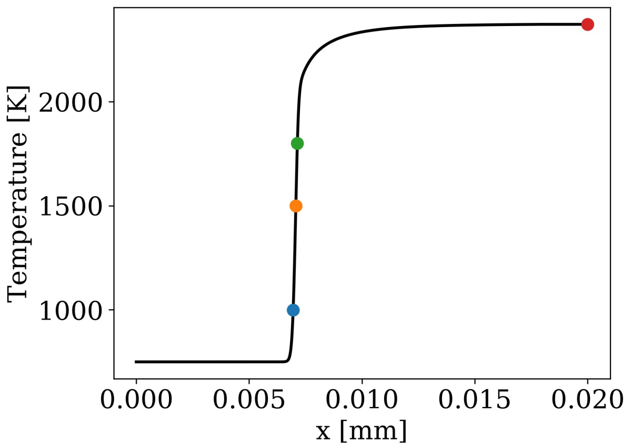


Probabilistic PCA via Autoencoders and Variational Inference

Methodology applied on a 1D flame in a canonical configuration:

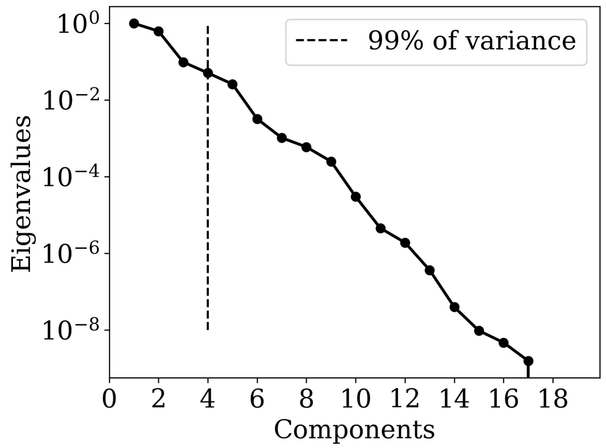
- Hydrogen fuel with different compositions
- Chemical reactions captured with 19 chemical species
- Flame is in steady-state

Spatial distribution of temperature of the flame



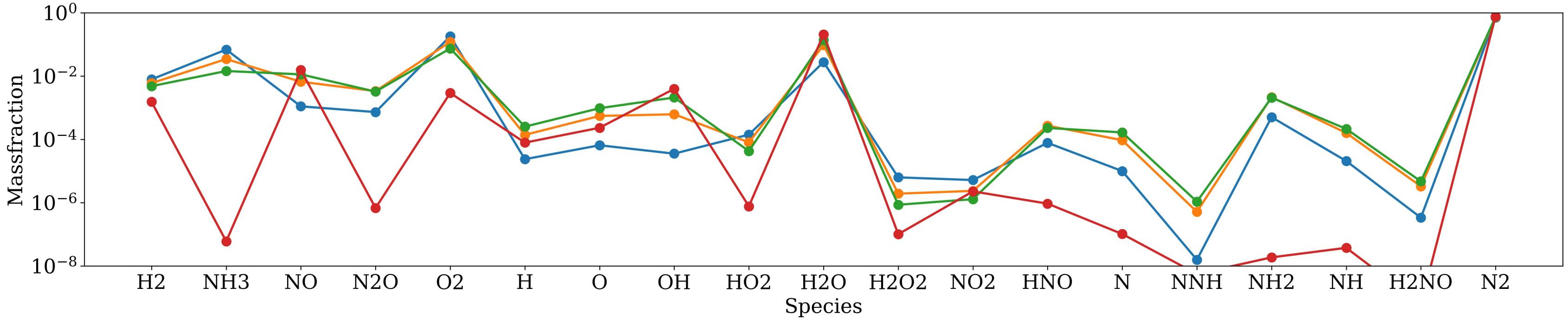
PCA

Normalized PCA eigenvalues



4 components required to capture 99% of variance

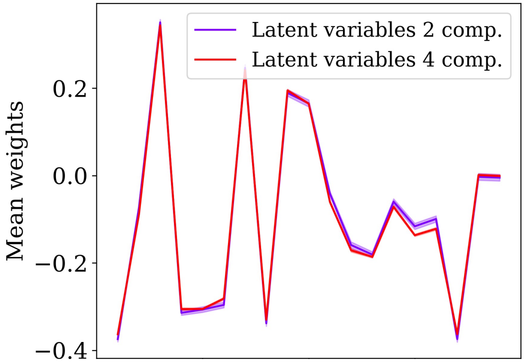
Composition realizations (snapshots) at varying locations



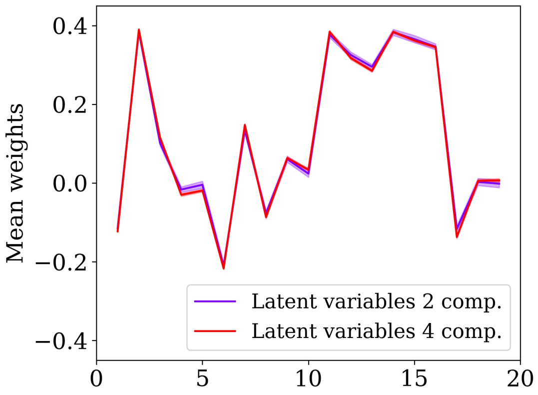
Sandia National Laboratories

Results of applying variational inference to train a probabilistic autoencoder using 1D flame data

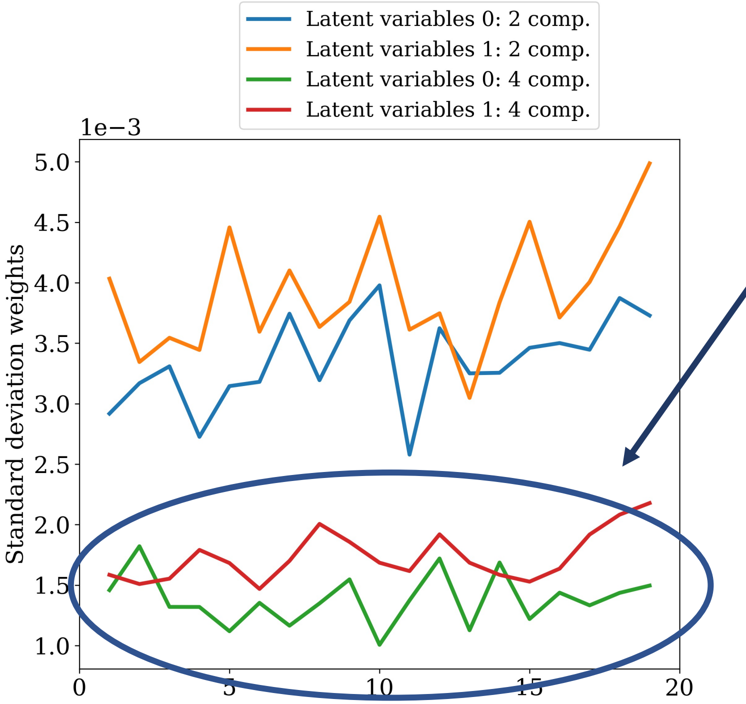
- Reduce dimensionality of composition space from 19 down to 2 or 4
- Utilize a mean-field approximation (with Gaussian PDFs) of the weight distributions
- Jointly infer observational/measurement noise intensity (assuming white Gaussian noise)
- Modeling achieved using Pyro's SVI capabilities with approx. 700 snapshots
- Mean variational parameters for the weights are warm-started using PCA (pre-processing step)



Weights, first latent variable

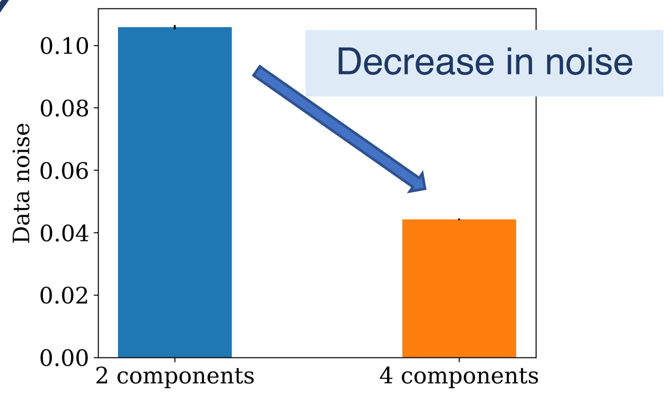


Weights, second latent variable



Uncertainties in weights

Decrease in standard deviation with 4 components



Observational noise estimates

Summary and future work

- PCA to expedite DNS of turbulent combustion
  - Dimensionality reduction of state space manifold
  - Limited training data
- New TL methodologies to reduce the number of snapshots required to train a PCA model
  1. Reducing the validation/testing errors of such models by leveraging data from similar domains
  2. Propagating parametric, model-form, and data uncertainties towards predictions
  3. Allowing for optimal ML model selection within the TL paradigm
  4. Safeguarding against negative learning (decreased accuracy due to task disparity, w.r.t. baseline)
  5. Consisting of strictly non-intrusive methods, applicable to most ML models without needing to modify the model (architecture) or implementation.
- Autoencoders used in lieu of PCA
  - Allows training using variational inference with established TPLs

Future work

- Application of presented TL framework to the training of probabilistic autoencoders
- Test framework on flames with different initial compositions